DOCUMENT RESUME

ED 203 729 HE 013 710

AUTHOR	Scriven, Michael
TITLE	The Evaluation of College Teaching. Professional Development Occasional Paper Number 3.
INSTITUTION	National Council of States on Inservice Education, Syracuse, N.Y.
SPONS AGENCY	Department of Education, Washington, D.C.
PUB DATE	[80]
NOTE	22p.
AVAILABLE FROM	National Dissemination Center, Syracuse University School of Education, 123 Huntington Hall, Syracuse, NY 13210.
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	*College Faculty; *College Instruction; *Evaluation Criteria; *Evaluation Methods; Faculty Development; *Faculty Evaluation; Grading; Higher Education; Student Evaluation of Teacher Performance; *Teacher Effectiveness; Testing

ABSTRACT

Perspectives on the evaluation of college teaching are considered. Some of the conditions that are needed generally include the following: the existence of an evaluation system that has a process for evaluating research and service, a clear definition of the difference between the evaluation of worth and the evaluation of merit, and a separate support system to assist faculty in improvement efforts. The following are not recommended for evaluating teaching: classroom visits by colleagues, focusing only on measuring the amount learned, and checking on the relative number of further courses in the same subject matter in which the instructor's former students enroll. It is suggested that the key component in the evaluation process is the student questionnaire, which should allow for an overall judgment of the merit of the instructor as an instructor. Other possible questions concern ethical or professional obligations of the instructor and components of instruction, such as ratings of the text, quizzes, and the grading system. It is suggested that instructional materials, tests, student grading, and the student's role in grading should be examined in terms of comprehensiveness, correctness, and currency. Other ways that teaching can be evaluated include assessment of actual learning gains and faculty self-development and self-improvement efforts. The use of faculty evaluation in a faculty development process, integrating and disseminating the results of faculty evaluations, and the key ethical/legal constraints on data gathering are also addressed. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED203729

Professional Development

Occasional Paper

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

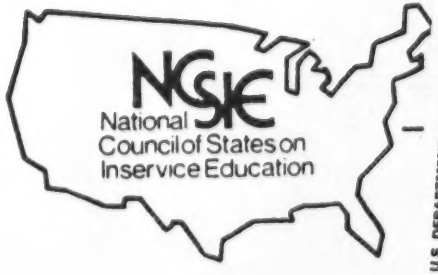
Michael Scriven

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

The Evaluation of College Teaching

Michael Scriven

HE 013 718



U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

The Evaluation of College Teaching

NATIONAL COUNCIL OF STATES ON INSERVICE EDUCATION

COUNCIL CHAIRS

Ronald Hall ..... *Chairman*
Louis Morelli ..... *Past Chairman*

PROJECT OFFICIALS

James F. Collins ..... *Project Director Syracuse University*
Carol Lewis ..... *Assistant Director*
R. Neal Appleby ..... *Administrative Assistant, Syracuse University*
Theodore Andrews ..... *Program Coordinator, Newsletter Editor*
Helen Hartle ... *Interstate Certification Project*
Diane Jones .... *NCSIE Project Officer, Teacher Corps*
James Steffensen ... *Associate Director, Teacher Corps Development Branch*
John Minor ... *Director, Teacher Corps*

MEMBERS

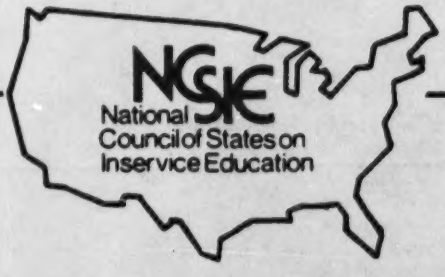
Organization	Representative
AACTE	David Imig
AASA	Robert Snyder
AFT	Pat Weiler
ATE	Robert Stevenson
NASDTEC	James Nohara
NEA	Roy Edelfelt
Teacher Centers	Allen Schmieder

State	Representative
Alabama	Billy Campbell*
Alaska	J. Kelly Tonsmeire
Arizona	Fred Sughrue
Arkansas	Austin Hanner
California	Richard Mastain
	William Webster
Colorado	M.D. Spurlin
Connecticut	Sigmund Abeles
District of Columbia	Joan Brown
Delawate	Ervin Marsh
Florida	Louis Morelli*
Georgia	William Leach
Hawaii	Eugene Yamamoto
Idaho	A.D. Luke
Illinois	William Gillies*
Indiana	Ronald A. Boyd
Iowa	Orrin Nearhoof
Kansas	Mary Martin*
Kentucky	Shirley Duff
Maine	David C. Tilton
Maryland	Irving Herrick
Massachusetts	Thomas O'Connor
Michigan	Paula Brictson*
Minnesota	Judith Wain
Mississippi	Russell Crider
Missouri	P.J. Newell
Montana	John Voorhis
Nebraska	Gerald Sughroue*
Nevada	LaMar LeFevre
New Hampshire	Gary Barker
New Jersey	Daniel Kunz
New Mexico	Alan Morgan
New York	Vincent Gazzetta*
North Carolina	Elizabeth Koontz
North Dakota	Lowell Jensen
Ohio	Nancy Eberhart
Oklahoma	John Folks
Oregon	Richard S. Jones
Pennsylvania	Ronald Hall*
Rhode Island	Edward L. Dambruch
South Dakota	James Hansen
Tennessee	John Gaines
Texas	James L. Kidd
Utah	LaMar Allred
Vermont	Henry Bissex
Virginia	Everett B. Howerton
Washington	Lillian Cady
West Virginia	Nicholas Hobar*
Wisconsin	Robert Skeway
Wyoming	Thomas L. Morris

*Member of the Executive Board

The Evaluation of College Teaching

Michael Scriven
Director, Evaluation Institute
University of San Francisco



The preparation of this publication was supported in part by funds made available by the United States Department of Education through Teacher Corps to Syracuse University Office of Sponsored Programs, National Council of States on Inservice Educçation Project. However, the opinions expressed herein are those of the author and are not to be considered as representing the opinions of the United States Government, the National Council of States on Inservice Education, any of the States which are members of the Council or of Syracuse University School of Education.

distributed by:
National Dissemination Center
Syracuse University School of Education
123 Huntington Hall
Syracuse, N.Y. 13210

Library of Congress Cataloging in Publication Data
Scriven, Michael.
The evaluation of college teaching.
(Professional development occasional paper; 3)
Includes bibliographical references.
1. College teachers, Rating of — United States.
I. National Council of States on Inservice Education. II. Title. III. Series.
LB2333.S37 378'.1222 80-26644

The Evaluation Of College Teaching

Michael Scriven

I. Introduction

The normal procedures for evaluating college teaching are so shoddy at the intellectual and the practical level, that it is hardly surprising that teaching is rarely rewarded in an appropriate way. That these procedures are intellectually sloppy is a reflection on both the scholars and the administrators of the academy, and that their methods of application are also slipshod tells us something more about the true value system of the academy which is not to its credit. Had colleges spent one percent of the time on this problem that they have spent on *internally*-funded research in the sciences or the humanities, we would now have something to be proud of instead of a source of shame. What follows has only the status of preliminary notes about the problems and their solution and easy ways to improve it greatly. I have tried to indicate the reasons why even these preliminary considerations clearly entail the rejection of essentially all existing systems.

II. Contextual Prerequisites

It is not only pointless but improper to proceed with a system for evaluating teaching that is supposed to operate well and ethically regardless of the administrative and legal context. The following are some of the conditions that must be present in that context.

A. The evaluation of teaching must be part of a *system* which also has an appropriate process for evaluating research and service, and a *specific* commitment to their *relative* weighting. Otherwise one is playing in a game for which only a few of the rules have been stated. It is notable that, for example, the University of Texas' system and the University of California at Berkeley's system have no such commitment to the relative worth of teaching. The fact that UCB requires—and enforces the requirement—that data on teaching, including student evaluations, must be included in a dossier before it will be considered for personnel decisions seems to show a valuing of teaching. It does not. The data may show that the teacher is a bad teacher, yet the "data requirement" does not result in a penalty; and if it shows/he is a good teacher, no determinate benefit results. *Specific relative weightings of all relevant criteria* of faculty assessment is the only non-vacuous procedure, and not incidentally the only equitable one. It is hardly surprising that in the two systems mentioned (and most others) the standards vary by campus and department.

B. You must have a system of administrator evaluation in place in order to avoid the entirely justifiable resistance of the "serfs" to being evaluated by the "folk in the castle" who are above such things themselves. Administrator evaluation, as it is commonly done at universities and colleges, is a truly remarkable phenomenon in that it succeeds in being even worse than the evaluation of faculty, and it occurs in even fewer cases.

C. The institution must have clearly understood and defined the difference between the evaluation of worth and the evaluation of merit, with respect to the evaluation of teaching in particular. (The distinction also applies to the evaluation of research and service.) There are a half dozen areas where the impact of this distinction is crucial, but I'll just mention two. Suppose that you have a teacher whose teaching is superlative by every reasonable standard, and who is also producing substantial quantities of absolutely first-rate research, while rendering impressive service to the profession, the campus and the community. If you are not, as an institution, absolutely clear that in spite of all this you may have to deny tenure to the teacher, you do not understand the difference between worth and merit. The decision whether to make an initial ap-

pointment, the decision on granting tenure, and the decision on pushing for early retirement by the various means available for that, should all depend upon the *worth of a faculty member to the institution* and not merely upon the faculty member's *professional merit*. The worth to the institution is essentially connected with such issues as the income generated by the student enrollment that the instructor produces, the special services to the institution's general mission that are provided by the instructor (e.g. a uniquely talented Latin scholar or teacher at a Jesuit institution, a woman mathematician in an otherwise all-male department that is anxious to recruit women as mathematics majors etc). Neither private nor public institutions can today afford to be awarding tenure on merit alone as if the enrollments were irrelevant, but many of them are locked into a system of faculty evaluation which makes it impossible to deviate from merit considerations even in extreme cases (except at the appointment decision). If worth is to be given any weighting at all, as it should be, then the exact size and limit of this weighting must be as carefully defined as possible in order that one does not get into obvious abuses such as firing faculty members whose political activities lead to the loss of large donations by alums or legislatures, the sort of abuse which results from thinking that "worth" means financial worth. It's not that worth doesn't have anything to do with financial worth (as in the enrollment example), it's just that it also has something to do with the integrity of the mission of the school and there is no worth left in an institution that abandons appropriate professional standards in higher education in order to achieve financial worth.

D. There *should* be an independent support system of some kind available to faculty to assist them in the effort to improve, so that the system of faculty evaluation is not merely punitive, and not *seen* as merely punitive (in which case it is unlikely to be viable). The system of independent support for improvement must be independent of the department chairs, the deans and the peers in order to avoid the disincentive to use it that would result if instructors know that administrators will know or can find out when it is being used by a particular instructor. Hence the consultations must be absolutely confidential as well as professionally based. Finding somebody who can provide the appropriate kind of assistance is very difficult because there aren't many who might be said to be qualified in the first place, and a large slice of *that* group is much too inclined to think that there is one true solution to the problem of how best to teach. The cost of providing this support system is low. One professional plus one secretary/assistant per 10,000 students can do this and *also* a number of other jobs that should be done on any campus such as keeping up with the research and practice in the field of faculty evaluation and teaching research, maintaining a small consulting library for faculty, improving the student questionnaire and managing its administration and data synthesis phases. This involves research-reading competence; but the helping role that works with the prima donna type of faculty is not something that prima donna types of teaching researchers are necessarily used to, so a nice balance of skills is required. It would be foolhardy to make appointments to this job for more than a trial period at first.

E. The background system must be one which has *consistent and appropriate practices*, not just consistent and appropriate rhetoric. A typical example of how not to achieve consistent practice is to have a system in which the quality of teaching is said to be important (which involves, amongst other things, the quality of the assessment of student work) while in fact departments are issued new positions and replacements for retiring—or departing—old appointments on the basis of enrollments. While this is admirable in the sense that it reflects an attention to a reasonable consideration that a decade or two ago was almost totally disregarded, it is absurd in that in many contexts it amounts to rewarding departments for—amongst other unattractive things—easy grading, which is inconsistent with the rhetoric of respect for high standards. Another example is the use of a student questionnaire which attends to the extent to which instructors get to know the names and personalities of individual students, provide ample office hours etc., all of

which are results easily achieved for small classes but not for large classes; and then to complain about declining enrollments, which this approach reinforces. Consistent systems are hard to set up, but the inconsistencies in systems are extremely expensive.

F. A reasonable modus vivendi with student government must be worked out for mutual benefit from student evaluation. (See section VI of this paper.)

III. The Definition of Good Teaching

The best teaching is not that which produces the most learning. No definitions in the literature avoid obvious counter-examples. The following definition avoids the so-far-identified counter-examples, but possibly has some of its own. The ideal teacher is one who has the maximal possible influence towards beneficial learning on the part of his or her students, subject to three conditions: 1) the teaching process used is ethical, 2) the curriculum coverage and process is consistent with what has been promised, 3) the teaching process and its foreseeable effects are consistent with appropriate institutional and professional goals and obligations.

Perhaps we can best illustrate what is intended by the three qualifications with some examples of what they exclude. 1) Unethical processes including not only cruel and unusual punishment but also those processes which are completely non-generalizable e.g. putting so much pressure on the students that they abandon their work for other classes in order to do the work for this one. It is often argued on weak *a priori* grounds that the use of a reward system such as a token economy or grades is unethical; in fact, the *failure* to use such a "reward system" is both unethical and unprofessional. It is unethical because it fails to inform the students about their progress or competence and/or it fails to show that the institution cares about quality work as opposed to minimally competent work etc.; and it is unprofessional because it keeps important information about the student's progress from the student and the student's advisors, parents and potential employers. It also happens to provide a legitimate incentive for many students, and not to use it is thus poor pedagogy unless you have direct evidence for a better approach. (Grading can even be arranged so as to make inter-student competitiveness impossible, while retaining the feedback required in striving for excellence.) So, even if grading reduced learning, it would have to be done and done properly. But—properly managed—it should increase learning.

2) Certain contracts are made by instructors and institutions, both explicitly and implicitly, to which they frequently do not adhere. Such contracts are involved in promises as to what a course will cover, made either in a catalogue, the departmental handouts, the faculty handbook, the union contract, the class handouts or in the language of the opening presentation in class. Apart from the simple sin of misleading advertising, much of the hierarchical structure of sequential curricula has been made laughable by the failure of instructors to adhere to these standards. Maximizing learning cannot be given precedence over these obligations.

3) It would usually be appropriate, if the maximization of learning were the only obligation of an instructor, to adjust the level of instruction to the class average so that more people would be able to benefit from it. But there are institutional and professional commitments that transcend this commitment. For example, if one is faced with a class of medical students who will graduate at the end of this year, and if one's obligation is to instruct them in professional procedures that they will be practicing upon an unsuspecting public a year from now, and if it is clear that most of them are so far off the pace that no instruction within the time available can get those students up to a level of competence, then it is professionally obligatory to focus the instruction upon those few who *can* be brought up to the appropriate standard and flunk the rest if they don't make it. Other obligations that are important and must be considered very seriously here include the problems of providing compensatory justice for minorities and women (which may mean focusing more effort there than shows up in maximizing classroom learning), and providing knowledge that is going to be expected by the instructor of a higher level course etc.

Some assorted notes that bear on the preceding points include the following. First, the "learning" referred to in the definition can certainly cover attitudes, in the rare cases where it is possible to justify the claim that teaching a certain attitude is an appropriate part of an instructor's obligations. Teaching the scientific attitude might be a case in point, as might be teaching the motivation to learn. It must also be remembered that teaching can often be done and often is done to a very important extent outside the classroom; hence evaluating it must involve looking for and at this out-of-classroom teaching.

Perhaps the most important feature of this definition of teaching is that it does *not* identify good teaching with the production of good learning. Considerations such as those mentioned preempt such a naive equation (one that I rather fancied in my earlier thinking about this subject); but there is yet another type of example that should be mentioned. If you have a number of students in the class who, for various reasons not connected with one's own performance, are not working hard enough to keep up with the class, one should not be downgraded as a teacher for the failure of these students to learn. A teacher's task is only to provide the *best possible* environment, not to guarantee that the results will be effective no matter how little effort is made by the students. It follows that one must never use on a student questionnaire the question: "How much did you learn from this class (that you think was valuable to you)?" or cognates of that question. One must instead use questions such as "How well do you think the instructor taught the course?" It is partly for this reason that the answers to student questionnaires are not to be regarded as intrinsically inferior to studies of the learning gains of students in class. They are *more valid* in one respect, in that they address the correct question.

A second incorrect definition of good teaching which has some supporters involves the identification of teaching with the transfer of learning from the teacher to the student. But of course—as has often been pointed out—inspiring the student to seek learning elsewhere may be the best approach to maximizing learning even in the short run and more likely in the long run.

IV. How Not to Evaluate Teaching

A. It will be clear from the preceding discussion that any way of evaluating teaching that *simply* consists in measuring the amount learned is oversimplified and can be extremely unfair to some teachers. It's also clear that evaluating a teacher's worth only or merit only will on occasions be entirely inappropriate. Let us turn quickly to a number of other errors that have been widely incorporated into quasi-systems for evaluating teachers.

B. Classroom Visits. Using classroom visits by colleagues (or administrators or "experts") to evaluate teaching is not just incorrect, it is a disgrace. Some of the reasons for this conclusion follow. First, the visit itself alters the teaching, so that the visitor is not looking at a random sample. Second, the number of visits is too small to be an accurate sample from which to generalize, even if it were a random sample. Third, the visitor is not devoid of independent personal prejudices in favor of or against the teacher, arising from the fact that the visitor is normally an administrator or colleague of the teacher and in his/her other role is involved in adversary proceedings, alliances, etc. with the teachers. Fourth, even if none of the preceding conditions make the whole affair ridiculous, which each of them does independently, there is nothing that could be observed in the classroom (apart from the most bizarre special cases) which can possibly be used as a basis for an inference about the merit of the teaching. That this is so follows inexorably from the results of the enormous number of studies on style research that have been done and summarized on various occasions, the sum total of these resulting in the conclusion that there are no style indicators that can be said to reliably correlate with learning now or in the future by the students in the classroom. Fifth, regardless of the fact that no observations of teaching style can legitimately be used as a basis for inference to the merit of the teaching, the visitor normally believes the contrary. This is often because the visitor has his or her own preferences as to a certain style, or has many years of expe-

rience in teaching this same type of course, and consequently believes that not doing it this way—or in one or two other ways that are approved—is doing it badly. These prejudices are totally without foundation, to the best of our knowledge, and should not be allowed to come into the evaluation of teaching at all. (Among the rare exceptions are the possibility that what the teacher is saying is known to be false by the visitor, *and* so grossly false as to constitute an impossible vehicle for teaching the truth; that the visitor observes racist or sexist or other immoral practices by the teacher; or that the visitor observes a total lack of classroom discipline to a degree which cannot possibly be reconciled with the continuance of a learning process of any kind. Since none of these events has ever been recorded on any of the classroom visiting sheets of the many thousands that I have either inspected directly or of which I have seen summaries, this cannot be taken seriously as a reason for making classroom visits part of teacher evaluation.)

Ultimately, the problem about the visitor is the lack of similarity between the visitor's head and the student's head. They are generally separated by several decades in their learning maturity, thus they may have substantially different vocabularies and conceptual repertoires, and they certainly lack much cultural similarity; because of this, the visitor is not a good indicator of how much learning is going on in the heads of the students. Since the secondary indicators of teaching style (i.e. everything besides empathy) turn out to be invalid, this leaves the visitor—or *should* leave the visitor—up the proverbial gum tree. I need hardly stress the fact that in spite of these staggering objections, the method of classroom visitation is the universal method whereby teachers in the elementary and high school are evaluated, and is—depressingly enough—being quite steadily implemented at the post-secondary level, on the grounds that it represents an improvement. No variation of this is of any value whatsoever; visits by experts are no better, visits by peers are no better than visits by administrators, *with respect to the personnel evaluation process.* There is a very limited role in which visits by a consultant are defensible in the effort to provide help to improve teaching *if it is already known that the teaching is very unsuccessful.* This is simply because the consultant may be able to suggest some options—not *optimal* alternatives, but just *possible* ones. But it's usually unnecessary to have the (costly) classroom visit—a tape recording or even a verbal report by the teacher plus the student evaluation forms is more than enough basis for consulting recommendations in most cases.

C. Teaching is usually evaluated without any serious attempt at evaluating the quality of the content of the course. This is the "methods-madness" that has made schools of education a laughingstock in the intellectual community and increasingly in the total community, for many years. From what we have said about the definition of good teaching, it is clear that one cannot make one's judgment of teaching merit entirely on the basis of the content of what is learned by the students, but to do so is much better than making the judgment on the basis of merely inspecting what is presented to students. Both should be considered, but with an emphasis upon student performance. Here is the one place where peer evaluation of a limited kind is appropriate—but evaluation of materials not process—and even here it is better to use people from another institution (but in the same subject-matter area), eliminating costs by trading services in kind with that institution. Such an arrangement—with its social pressure—tends to do much more to improve standards than in-house materials evaluation does anyway, apart from increased validity.

D. Another popularly acclaimed though infrequently employed measure of teaching success is checking on the relative number of further courses in the same subject matter in which "graduates" of the instructor being evaluated enroll. This is an unethical indicator and its use cannot be countenanced by central administration because it is non-generalizable. That is, one can only score points on this dimension by stealing, buying or seducing students from other departments. It's exactly like the process of getting more work out of the students by having

them give up on their homework for other courses. It's a great indicator if you treat the welfare of the discipline as the *summum bonum*, rather than the welfare of the institution or of higher education or of the humanities or of the society; or of the student (with a minor qualification).

E. Most student questionnaires are improper bases for faculty evaluation, because they generally involve ratings on style or ratings on non-generalizable indexes etc. But the way in which the data from them is synthesized is also a problem. If you only use mean ratings, then you overlook the important case of the instructors who are tremendously successful with a sub-group of the class. Although their mean score may be no different from that of another instructor who receives weak ratings from everybody in the class, their teaching potential is obviously different and a sensitive administration should be making some provisions to tap such a promising source of inspiration in a more appropriate situation than the particular class that resulted in a low overall score. Or the consultant may be able to work out a way to generalize from this promising basis to the other students. Similarly, averaging the means from courses taught at different levels may be unfair to good teachers of introductory (or graduate) courses.

There are a number of other ways not to do the job that are widely used, but let's turn our attention to at least a preliminary sketch of how the job might be done properly.

V. How Teaching Should Be Evaluated

A. The key component in the evaluation process is the student questionnaire, but the piece of paper itself is only part of the story. The support system has got to provide methods for administering this that are bullet-proof against complaints about the possibility of selective return and prompting. A good straightforward approach is to have the central administration staff (a cheaper fallback system is using the departmental secretaries) take the questionnaires out to each class, have the professor leave the room for a few minutes, provide a brief explanation of the process and how the results are to be used (possibly in writing), encourage questions, and pass out questionnaires to be filled out by every member of the class. One should get a 99 percent return rate from those present. Where knowing that the questionnaire will be distributed on that date is an incentive, the date for distribution should be set in advance if one wants most of those enrolled in the class to attend. (This does weaken one's defenses against certain types of preparation of the students by the professor, but this can be controlled by asking about it on the form so it is a less worrying problem than low attendance rates.) "Absentee ballots" are not a good idea, logistically or for inferential purposes. One can get that kind of response rate, and tolerance from the faculty for taking time out of their classes, if the whole process takes less than about 5 minutes. This is the first of two compelling reasons for using a very short questionnaire.

The second is that there is not the slightest justification for using a long one. The usual questionnaires with twenty to seventy items on them are fishing expeditions of an entirely improper kind. Unless you can say of a particular question that *it is demonstrable that an answer of a particular kind indicates merit or demerit in teaching*, then it has no place on the personnel evaluation questionnaire. Other and more detailed questions may be used in a questionnaire designed by the instructor with the assistance of the consultant (or perhaps the other way around), when an instructor wants to get specific feedback on some specific effort or style venture in which he or she is interested. But no such commitments can be legitimated for general evaluative purposes and no such questions should ever be on a form which is seen by anybody except the instructors and consultants of their choice. If placed on a form along with legitimate questions they will be likely to bias the response of somebody who favors or disfavors the style that they uncover, and such biases are illegitimate. (They would be certain to seriously affect the legitimacy of any personnel decisions that were appealed in court, for example; but the ethical point is more serious than that.)

Since one cannot help but be worried by the contamination of instructor ratings by *irrelevant personality factors* and by factors connected with *like or dislike of the course*

itself, one should make some effort to siphon off those other considerations. A simple way is to stress the contrast between these considerations in the introduction to the basic judgmental question on the questionnaire, i.e. between liking and disliking the instructor as a person and thinking well of him as an instructor; and between liking and disliking the course content (of a required course) and thinking that the instructor did or did not do a good job of teaching it. Another approach to siphoning would consist in asking specifically for 1) evaluation of the instructor as a person, and 2) of the course itself, and *then* asking for 3) an evaluation of the job done by the instructor in teaching this course. The course evaluation could then be torn off the form and sent to the chair and/or dean for course evaluation; the personal evaluation could be torn off and thrown away or given to the instructor if she or he requested it; and one would then use the rest of the form for personnel evaluation.

The crucial question in the rest of the form is a request for the overall judgment of the merit of the instructor as an instructor. There *are* some other possible questions on such a form, of a kind not often encountered. They should come before the requests for an overall rating since they are likely to depress it, (and increase its validity) which is fine because the main problem with the usual results is that the scores are too high to allow adequate headroom in which exceptionally good teaching can distinguish itself. These other (possible) questions are of two kinds. The first concerns matters that can be described as ethical or professional obligations of the instructor and can be phrased in either a positive or a negative way, or—better—alternately. They may concern such matters as the match between the preannounced content of the course and the actual content; the match between the content of the course and the content of the examinations; the extent to which the reasons for grades were explicitly and adequately justified; the extent to which the possibility of appealing a grade was explained; the frequency with which scheduled class and office hours were met; the use of racist or sexist or other bigoted remarks or materials and so on. This is the so-called "Black Marks list" (when phrased negatively) and it is surprising the extent to which it reminds instructors as well as students of the minimum professional obligations of the instructor, obligations that can rather easily be discharged and which one constantly finds are an underlying cause of dissatisfaction and bad holistic ratings.

The second kind of legitimate question consists in a listing of the components of instruction for independent "micro-assessment" i.e. "please rate the catalog description/text/quizzes/grading system/ class handouts/sections or labs or fieldwork etc.". I now feel that a form with these component-evaluation questions plus the macro-question, plus an optional request for suggested improvements in the form and process of administering (using it) is the best basic form, though one can reduce the effort by using only the macro-evaluation, going to component analysis only when the instructor requests it or is seriously and regularly below the mean for comparable courses. At that point, where remedial action is necessary the components evaluation is an obvious help. When a crucial personnel decision has to be made, I think that the "professionality" questionnaire provides a useful and valid supplement to information from the other two. Such data should be accompanied by comments on it from the instructor.

B. There must be careful examination of the quality and professionality of content and process; the three qualities here are currency, correctness, and comprehensiveness. Ratings must be made on the basis of a sample of 1) the materials provided, 2) the texts required and recommended, 3) the exams, 4) the term paper topics, 5) the student performances on the preceding, 6) the instructor's performance in grading student work, 7) the instructor's performance in justifying the grade and providing other helpful feedback.

C. Actual learning gains. These are only going to be useful if one can in advance specify and justify the comparisons that will be used. By themselves, they have no legitimate interpretation except, perhaps, where they are negative. In cases of multiply taught sections for an instructional course, the comparison between sections can be extremely useful. In cases where there are

national norms they can sometimes be useful. And with respect to instructional improvement, one can run a comparison against one's own previous performance and discover whether slight or large changes in approach or text turn out to have significant results.

The main aim of the comparisons involved is to try to determine what could reasonably have been taught to the students; it's always fairly straightforward to find out what was actually taught, but that is evaluatively useless without some sense of what was possible. If allocation of students between the afternoon sections of a large intro class is random, then one can indeed discover something quite important about this, if one is prepared to look for patterns that extend across a couple of years. Shorter periods of study are likely to yield results due to unrecognized idiosyncracies of the particular classroom, time, groups present etc.

D. Professional Development Dossier (Self-evaluation and Self-improvement information.) Although the results of self-improvement should show up eventually in improved performance on the above scales, it's worth including it directly since doing so encourages faculty efforts in this direction, hence speed up the improvement; and it also improves the validity of decisions that have to be made under a time constraint. Under this heading we expect the submission of a rationale for each course's methods and coverage (where the latter is the instructor's option), evidence about professional development activity such as readings, courses, workshops, consulting aimed to improve teaching, and most importantly, a description of planned and performed experiments on the individual instructor's own teaching approach. A basic list of courses taught, environments, and grades; and of committees (instructional, departmental, school and college) should be included here preferably as hard-copy output from the institutional data base.

The preceding list of four dimensions covers the four crucial *types* of data that should be gathered. There are ways in which these can be ramified usefully and without significant extra cost. For example, feedback from the students should be supplemented by feedback from teaching assistants or teaching aides where these are employed, and where confidentiality of the responses can be preserved (as is usually the case). I have sometimes previously included a separate ethical component or a "justice dimension" (as I then called it), but I have here amalgamated it into "professional content and process."

For faculty self-development purposes, the questionnaire can of course be expanded to include an open-end question in which there is a call for identifying particularly good and particularly bad features of the instruction, but I usually keep these responses down in volume (because it's so hard to simplify large numbers of these responses) by requesting that they be provided only when either the top, or the bottom two scoring categories is used. (i.e. only when an A or a D or an F rating of the instructor is given by the student). This also improves their utility, because it avoids the probability of interpreting e.g. favorable responses with long lists of criticisms attached.

In addition to the ratings of the instructor done by students during the class, there is another type of student rating that is exceptionally valuable, at least when done sporadically, and that is the "exit" interview done at graduation. It falls midway between the in-class rating and the alum rating, and is vastly better than the alum rating because the rate of return can be around 100 percent instead of around 30 percent (rendering the latter essentially useless for personnel decisions), memories are more reliable at that point, it refers to a more recent version of the courses, and causal inferences are more likely to be valid at that point. Exit interviews have a quite disproportionately large proportion of the students, but who do not show up in any other search.

Notice that the approach outlined here does a great deal to protect faculty from two sources of injustice that are particularly pervasive in current systems. The first is the kind of injustice that makes it impossible for the teacher of a generally unpopular but absolutely essential prerequisite course e.g. calculus for architect students, to score really well compared to the norms against

which she or he will usually be judged. We make allowances in the above system for at least three ways in which such an instructor will be able to emerge as first-rate. The other kind of injustice relates to the individuals who put an enormous amount of work into developing a course, keeping it up to date, and making it as effective a teaching device as they possibly can, *at the expense of* lots of happy friendly humanistic interchanges and socializing. When an evaluation system is issued which does not pay a great deal of attention to the content of the course, and which throws in irrelevant requests to have students rate the course on "touchy-feely" dimensions, these efforts are largely unrewarded.

Although peers—one kind of expert—must be used for evaluation content, they are usually not competent to handle all that falls under B. although they should be. For there is one kind of usually hidden behavior that should be picked up by an evaluation system, and which does require a different kind of expertise to assess, and this is the technical competence of the instructor in pedagogy e.g. in writing reasonable examination questions that are unambiguous, involve adequate coverage of the intended target area, are not overcued, etc. This kind of quality is rather easily picked up by somebody with good skills in test construction, and such a person should routinely review samples of the tests of those instructors that are up for (or a year from) particularly important personnel decisions like tenure.

Another point at which such experts have some relevance concerns the way in which exams are graded or marked. There are professionally required standards here, with which virtually no faculty members at universities have the slightest familiarity. As a remedy for this, administrators should request that as a normal part of the process of talking about self improvement, the instructor fill out a form indicating how papers are in fact graded. The kind of grading that is legitimate involves at least the following requirements; the papers are graded "blind" e.g. by having the students write their names inside the back cover of a blue book or in any of a dozen other more foolproof ways; the papers are graded question by question, not paper by paper (this is to avoid the known large halo effect that results from having read a good, or bad first question by the same author just before reading the second question). When one has graded all the nth questions in a given set of papers one must then go back and grade the first six to ten papers again in order to see whether one's grading standards have slipped, upwards or downwards (as they often do); the pile of papers should be shuffled after each question is graded so that different students stand the brunt of one's fatigue (or initial optimism) for different questions; and so on. With multiple choice exams, many of the preceding requirements are otiose; but the requirement for technical proficiency in posing them becomes enormously more important and is something that is beyond the competence of nearly all academic instructors at the moment. (It's illuminating to talk to people at the Educational Testing Service about the difficulty of finding individuals who can write good multiple choice items. I think one can tie in these practices with those required by the "Black Marks List" and call them the Professionality Dimension (E). But they can also be left buried in A and B.

Much of the preceding refers to summative evaluation for personnel decisions. For that purpose, it is quite important to give out the class questionnaires at about the same time in the term for all instructors. What is the best time? Faculty at USF argue that the tenth through the twelfth weeks of class (semester system) would be about right; not too near to the beginning to give a reading on the basis of inadequate experience, not too near to the end so as to interfere with the intensive review period for the examinations and the occasional drop in attendance while students stay at home. Getting all the classes visited within this period is logistically feasible for a relatively small campus, but would involve substantial difficulties in a large one. But validity considerations suggest that administration of questionnaires after final grades have been seen is preferable and this also increases headroom on the distribution. See section VI for more on this approach.

One could turn to the use of senior departmental secretaries to distribute and col-

lect questionnaires, but there is still a credibility problem about that. It is essential not to use graduate students or students from the same campus, for similar reasons. The system of having each class elect a student who will do the job and take the materials over to the central administrative offices has worked quite well at some institutions. But it does not provide a person who can answer questions and exhibit inside knowledge about the importance of the process and its results. Under no circumstances can one go to mail ballots or "drop it in the nearest collection box" ballots, because the response rate deteriorates seriously, and it's hard to justify validity with under 85% return rates especially if variable. One must select the best procedure in terms of the accessible resources at a particular site.

Similarly, although it is rather easy to use optically scannable cards, this gets one into the business of providing the appropriate marking pencils, which is to be avoided at all costs. It's better to use straightforward forms and have them keypunched; it turns out that this approach—rather surprisingly—is less expensive. The computer program for combining the data on the type of questions that we have been talking about is of course extremely simple and can provide a variety of interesting readouts. After investigating variations in small and large classes, required and optional classes, upper and lower division classes and differences between fields, one usually finds that most of these differences are not sufficiently important to be worth reporting separately, although one can run them through the machine for a check every time the system is operating. One should also have the computer automatically flag performances that are either a standard deviation above or one and two standard deviations below the averages, not because a standard deviation means something in terms of traditional significance with the kind of skewed distribution you get here, but because it's a convenient and quite appropriate flag. If one does not use a set of preliminary questions on components or professionality, one will probably have to set the upwards flag at half or two-thirds of a standard deviation in order to get any success stories at all.

How often should this process be run? In my view, every second course or every third course given, stratified by upper and lower division and graduate categories, is a good compromise. Moreover, it cuts the personnel and time requirements for distributing and collecting the forms by a half to two-thirds, besides giving faculty a chance to do something about an unsatisfactory readout before the next evaluation is upon them.

There are certain by-products of this activity that are of some interest to administrators. For example, one can get a very good readout on the actual teaching load of the faculty, in terms of numbers of classes and in terms of numbers of live bodies; it turns out that there are sometimes a startling number of "phantom" classes i.e. classes that are not in fact meeting although they haven't been cancelled, usually because there weren't enough people to justify a scheduled class meeting and the instructor didn't want to convey to the chair or dean the fact that her or his load had partly evaporated. A comparison of the number of grades awarded with the number present at the time of the administration's questionnaire will give an attendance rate ratio which is also something that bears watching for faculty members who are trying to improve their performance. The number present for the evaluation should not be printed out on the summary sheets without a reference to the number absent, if the registrar's computer can be brought to produce a current figure on that. Printing out some index of bi-modal distributions is also easy, but should in general be reserved for the situation where someone is looking for help.

VI. How to Use Faculty Evaluation in a Faculty Development Process

Most of the preceding refers fairly specifically to summative evaluation; it is time for us to say something more specific about formative evaluation. In general, there is no need at all for this to involve the kind of rigorous supervision of questionnaire administration we've been talking about to date. Nor need it be done at any particular scheduled time. In my classes I have frequently used the following eccentric procedure. Evaluation forms are distributed to

everybody that comes into the class the moment they walk in the door on the first or second or third day while the tourists are still shopping around. The forms request the student to turn them in by putting them on the front desk, or by giving them to another student to turn in, if they decide not to continue with the course after even one or two sessions. The reason for this is that one wants to know whether there was something misleading about the advertising, needs-irrelevant about the content or off putting about the early presentations, and this is the best possible way to pick is up. Perhaps there was something about the way that one was outlining the proposed examination and assignment process that seemed particularly formidable. One won't pick these objections up from people that left because of them and hence aren't there toward the end when the usual forms are distributed. Indeed, as you reflect upon the usual process, you see that it is heavily biased in favor of the instructor because of the highly self-selected population that stays around. One is losing much of the critical and negative feedback, and I personally find that I learn a good deal from it, though I won't pretend that I like reading all of it.

With respect to students who stay around, one next requests that all of them should feel free to submit anonymous suggestions and criticism at any time throughout the term. One needs to make very careful and specific arrangements about this, to preserve their anonymity and to encourage them to see that one is going to do something with the results: e.g. issue feedback forms and pass a collection box around once a week. After the midsemester, one requests that a full component analysis form be submitted with perhaps a half dozen questions on it related to how fair the exams and grades were, the coverage to date, the treatment of students raising questions in class etc. Now one has the chance to show that this input really is useful, by discussing it and taking whatever steps seem best to improve the class in the light of it. This is also a very good time to demonstrate that some of the criticisms are mutually contradictory and hence that it isn't easy to satisfy everybody. Next one applies to the faculty Senate or the appropriate administrator for permission to give an early final. The final is given during the last class (or two classes) and is corrected in time to be returned at the time originally scheduled for the official final in finals week. Attendance at the session is required, as it would be for a final, upon penalty of receiving an incomplete in the course, and possibly a grade penalty. At that last session, the exam questions are gone over in front of the class, an answer key with "model answers" is handed out, a sheet of actual and unsatisfactory answers that illustrate certain types of mistakes are handed out with the appropriate commentary on it, an opportunity is provided to raise objections to the grade with the instructor and perhaps with teaching assistants, and *then* the students are required to hand in a final evaluation form on the class. The students who had been optimistically supposing that they were going to get an A and, in the light of this, had given the instructor quite a good grade in the tenth, eleventh or twelfth week, are now suddenly confronted with reality and their evaluation of the instructor is often affected. Too bad—but one cannot expect them to provide as good a judgment of the course prior to receiving a grade for it as subsequently. Moreover this arrangement—apart from improving the evidential basis of the student evaluation and hence probably the validity of their ratings—makes the exams part of the learning process, not an arrow shot into the air and falling to earth one knows not where except for a grade. Improvement and effort should be directed by the most accurate evaluation possible, *whether or not the institution uses the best system.* That's the first point about formative evaluation (evaluation for development.)

In the course of formative evaluation, it is entirely appropriate to ask questions of the students about style, *where one is striving to achieve a particular style.* But it is even more appropriate (because of known validity) to ask some microquestions to identify problem areas such as the text, handouts, quizzes, mid-term assignments, grading process, treatment of questioners, availability in office and after class, and final exams. There are no style assumptions lying behind the belief that it is better to perform well in

each of these areas. Whereas it is certainly possible to have all instructors ask everybody about these matters, i.e. build them into the general form, the consequent increase in the load on the computer is not only costly but less effective because the rigid time for collection of the data does not provide one the opportunity to see whether improvements have been effectuated that seem to work well with the class. A one-shot feedback is never as useful formatively as a sequence of feedbacks.

If an instructor is getting bad ratings overall that do not seem to be explicable in terms of, for example, the peculiarity that the course is required for non-majors, then she or he should certainly go to the kind of components analysis form that I have just mentioned. Professional consulting at this point will often suggest a number of ways to improve performance. Only if all of that fails should one consider going to an analysis of style and a discussion of alternative possible approaches to *that*, because the results of style research make this a last resort and there are still other non-style reasons to worry about first, such as the giving of indiscriminate A's, which may better serve to precipitate a request or requirement that the instructor alter a particular teaching process because it's unprofessional.

I think that *most people* who have the requisite content knowledge are capable of becoming rather good instructors; but I don't think that many of them do so, and I don't think that *all* of the faculty—even those with tenure—either could or will become or remain reasonable instructors. Hence the evaluation system needs to be one which provides a solid basis for proceeding *first* to put pressure on an unsatisfactory instructor to upgrade performance, and *second* to remove them either from the faculty or from the instructional faculty if the first result is not possible. We are now moving away from the legal model of the "master-slave" relationship in colleges and towards the legal model of "just cause" for dismissal, as a result of the affirmative action legislation and the gradual development of an orientation toward collective bargaining. So previous fears about violations of academic freedom, which were the most important basis for tenure, are somewhat less of a worry and can certainly be taken care of via protective contracts, whether formal or informal. The residual infamy of tenure, due to the number of people who are "burned out" but kept on by it, stands as a proclamation of lack of responsibility that is becoming increasingly prominent as the hard times for higher education develop. We can't afford to continue that way, and we can't move any other way without a rock solid process for evaluating teaching—*and for improving it*, which must be the court of first resort.

VII. Integrating and Disseminating the Results of Faculty Evaluations

While the results of formative feedback go to the instructor alone, or to some consultant that is chosen by the instructor, the summative evaluations from students go to the responsible first level administrator. They must be combined with the information in the files, direct measurement of learning gains (if available) and appropriate comparisons, ratings by teaching assistants, the quality ratings made by topic experts of the content and by process experts of the examining and certain other professional processes. In addition, considerations of worth must be brought in for the appropriate decisions. The integrative process is a very tricky one and all the preceding work may be wasted if bias can slip in here as it usually can and does. Without getting into the minutiae of the topic, it is not easy to give a complete outline of what should happen at this point. But the ideal toward which one should strive is clear enough; the integrative process should be *a simple linear weighting and summing procedure*. In this respect it should be the same as the procedure for combining the results of the teaching evaluation with the results of research and service evaluations. Extreme rigidity in the process is far better (although not necessary) than allowing the department chair or the dean to do a holistic "seat of the pants" synthesis of the data at this point, a synthesis which is a principal way for bias to come in.

There are of course a variety of situations in which one does not want to use *equal* weighting for all these components, but it should be a matter of extreme openness how the actual weights are determined and

what they are; and where it is possible to arrange for individual variations in the "contract."

What we have sketched above is a rather incomplete version of what is only a *fairly* complicated procedure. It's a perfectly comprehensible procedure and can readily be made a rather equitable one. One of the results to be avoided at all costs is the pressure to go with different evaluative approaches by different departments or schools. It may be a useful political procedure to start off with that as a possibility in order to get people on board who would rather die than admit that their teaching process in a lab or a clinic has anything in common with the teaching process in a philosophy seminar or a basic statistics course, but the real truth of the matter is that any such view is complete nonsense. There is nothing about any of the procedures recommended here that will not work perfectly well for any course offered on a university campus, from the school of music to molecular biology. One may want to add an extra question or two here, drop one or two there, especially from the formative evaluation procedures, and perhaps even from the "Black Marks List." But these are very minor changes and the simplest way is to put all of the questions on all of the forms that are generally distributed, with a "not applicable" option for the students. The backup use of other forms, with reference to style, where necessary or desired, may certainly be made as idiosyncratic as the instructor pleases, and the departments may have a preferred version of these. But those forms should not get into the personnel review process *even at the departmental level,* without both legal and ethical expert opinion.

VII. The Key Ethical/Legal Constraint on Data-Gathering

The underlying reason for taking such a firm line about this point is not merely the absence of any scientific evidence that connects particular approaches to teaching with successful results; it is a much graver matter. The error in racism, or sexism, does not lie in the empirical falsehood of the claim e.g. that the crime rate amongst blacks is higher than it is amongst whites, or that management experience is much less likely to be present in a woman than it is in a man; both these and many other similar generalizations are true. But their truth fails to justify the appeal to them in order to make a personnel decision adverse to a black or woman candidate. And their truth does not justify such an appeal, not because it happens to be illegal, but because it happens to be scientifically indefensible. Generalizations like this are very much *less* reliable as predictors than inferences made from track records of the individual candidate. It is additionally and ethically unquestionable that any use of such generalizations would lead to self-fulfilling and socially undesirable results, so on that additional ground they must be abhorred. But one does not need the additional ground in order to see that, no matter what the results of research on teaching styles ever turns out to be, one will *never be able to use those results for making decisions about individual instructors because to do so would be a case of guilt by association.* (One may well ask what the *point* of research on teaching styles is, if it can never be used in this way. The best answer is that it can increase our repertoire of *promising options.*)

A final point on the ethical side. I do not think that it is ethical, and I suspect it may not even be legal, to deny the students who generate the key evaluation judgments in this whole process the opportunity to see the summarized results. One might certainly argue against it with respect to some detailed written-in remarks, because those cannot properly be used without balancing them against all the rest, and doing that in a systematic way is something that students may not be in a position to do; but the holistic responses and their distribution is something one cannot properly withhold from the students. In addition to questions of propriety, there is the further thoroughly unattractive possibility that the students, frustrated in their attempt to get these, will set up their own evaluation system and one will then *either* have to cooperate with it, which increases the chance of questionnaire overload with a reduction in response rate and validity, as well as a loss of class time; *or* take the extremely unattractive position that students are not allowed to poll their peers in

order to evaluate instructors.

Some compromises will be necessary because the kind of evaluation of instructors that students are interested in doing is as different from the kind that we have been talking about as formative is from summative. Students are interested in such questions as the assignment-mode, the cost of the textbooks and whether they are really necessary, and certain teaching style variables which can well be an appropriate basis for the student with a certain learning style to use in selecting in or out of a particular class. But compromises of this kind should be made and can be made in the interest of a campus-community approach. Student morale is not well served by leaving them out of this process, and in the long run faculty evaluation and faculty improvement suffers from not having input from these student points of view.

VII. Conclusion

Personnel evaluation in general is a difficult field and one that is usually done with shocking incompetence as one can readily discover by studying the forms used by the White House, or by large corporations or the Armed Services. If it is to be improved, presumably the academies should be the source of the leadership. It is they after all, who do most of the relevant research and interpretation. In the key matter of the evaluation of teaching at the college level they have shown a disgraceful disinclination to get into it, and in recent years when that distaste has partly evaporated, a disgraceful lack of capacity to do it in a defensible way. Perhaps the above remarks will prove sufficiently irritating to lead to improvement, even if they are not themselves accepted as suggesting appropriate improvement.

A Partial Bibliography Related to Clinical Supervision

Alfonso, Robert J.; Firth, Gerald R.; and Neville, Richard F. *Instructional Supervision: A Behavior System.* Boston: Allyn and Bacon, 1975.

Beegle, Charles W. and Brandt, Richard M. *Observational Methods in the Classroom.* Washington, D.C.: Association for Supervision and Curriculum Development, 1973.

Blumberg, Arthur and Cusick, Philip. "Supervisor-Teacher Interaction: An Analysis of Verbal Behavior." *Education,* Vol. 91, 1970.

Boyan, Norman J. and Copeland, Willis D. *Instructional Supervision Training Program.* Columbus, Ohio: Charles E. Merrill Publishing Company, 1978.

Brammer, Lawrence M. *The Helping Relationship: Process and Skills.* Englewood Cliffs, New Jersey: Prentice-Hall, 1973.

Cogan, Morris L. *Clinical Supervision.* Boston: Houghton Mifflin Company, 1973.

Goldhammer, Robert. *Clinical Supervision: Special Methods for Supervision of Teachers.* New York: Holt, Rinehart, and Winston, 1969.

Hyman, Ronald T. *School Administrator's Handbook of Teacher Supervision and Evaluation Methods.* Englewood Cliffs, New Jersey: Prentice-Hall, 1975.

Koran, J.J. "Supervision: An Attempt to Modify Behavior." *Educational Leadership,* Vol. 26, May, 1969.

Kysilka, Marcella L. "Research in Review/ Clinical Supervision: A Review of the Research." *Educational Leadership,* Vol. 35, April 1978.

Reavis, Charles. "Clinical Supervision: A Timely Approach." *Educational Leadership,* Vol. 33, February 1976.

The Supervisory Process: Helping Teachers to Improve Instruction - Videotape, Washington, D.C.: Association for Supervision and Curriculum Development, 1978.

Some Other Documents Providing Information Related to Teacher Evaluation

Evaluating Teacher Effectiveness—ERIC Abstracts. Eugene, Oregon: ERIC Clearinghouse on Educational Management, University of Oregon, 1971. *Evaluating Teacher Performance.* Arlington, Virginia: Educational Research Service, Inc. 1978. *Evaluating Teachers for Professional Growth: Current Trends in School Policies and Practices.* Arlington, Virginia: National School Public Relations Association, 1974. *Evaluating Teaching Performance.* Washington, D.C.: Educational Research Service, American Association of School Administrators, 1969.

NATIONAL DISSEMINATION CENTER
PROFESSIONAL DEVELOPMENT PUBLICATIONS

PUBLICATIONS ORDER FORM

No.	Title	Amt.
_____	Perspectives on Preservice and Inservice Education Louis Rubin	$4.00
_____	The Role of the State Education Agencies in Inservice Education Wendell Allen	$3.00
_____	Inservice Education: State Scenes National Council of States on Inservice Education	$6.50
_____	Sources & Resources: An Annotated Bibliography on Inservice Education National Council of States on Inservice Education	$8.50
_____	Inservice Education: Priority of the '80s Margo Johnson	$4.50
_____	Providing Leadership for Inservice Education Proceedings of the 1979 NCSIE National Workshop	$4.50
_____	State-Level Teacher Performance Evaluation Policies Lou M. Carey (NCSIE Occasional Paper Number 1)	$2.00
_____	Merging Resources of the IHE and the LEA for the Improvement of Inter-Institutional Programs Dwight W. Allen and The Competency Centered/Field Based Early Childhood Certification Project Gordon E. Eade (NCSIE Occasional Paper Number 2)	$2.00
_____	The Evaluation of College Teaching Michael Scriven (NCSIE Occasional Paper Number 3)	$2.00
_____	Expanding the Implications of Public Law 94-142 Maynard Reynolds (NCSIE Occasional Paper Number 4)	$2.00
_____	Issues in Inservice Education: State Action on Inservice Education National Council of States on Inservice Education	$4.00
_____	Involvement: A Study of Shared Governance Bruce R. Joyce	$4.00
_____	Toward Meeting the Needs of the Beginning Teacher Kenneth R. Howey and Richard H. Bents, Editors	$4.50
_____	**Set of Five ISTE Reports**	$20.00
	or	
_____	ISTE Report I—Issues to Face Bruce Joyce, Ken Howey, and Sam Yarger	$3.00
_____	ISTE Report II—Interviews: Perceptions of Professionals and Policy Makers, Bruce Joyce, Kathleen McNair, Richard Diaz, and Michael McKibben	$5.00
_____	ISTE Report III—The Literature on Inservice Teacher Education: An Analytic Review, Alexander Nicholson and Bruce Joyce	$4.00
_____	ISTE Report IV—Creative Authority and Collaboration: A Collection of Position Papers, Sam Yarger et. al.	$5.00
_____	ISTE Report V—Cultural Pluralism and Social Change Richard Brandt et. al.	$5.00
_____	Meeting the Special Needs of Students in Regular Classrooms James F. Collins and Joseph A. Mercurio, Editors	$3.00
_____	Criteria for Describing and Assessing Competency Based Programs J. Bruce Burke, John H. Hansen, W. Robert Houston, Charles Johnson	$4.00
_____	A General Catalog of Teaching Skills Richard Turner, Editor	$9.00
_____	A Catalogue of Concepts in the Pedagogical Domain of Teacher Education Bryce D. Hudgins, Project Director	$4.00
_____	Competency Assessment, Research and Evaluation W. Robert Houston, Editor	$3.00
_____	Closing the Knowledge Gap H. Del Schalock	$3.00

_____	Resources for Performance-Based Education & Supplement A W. Robert Houston	$3.00
_____	Social Studies Teacher Competencies Minnesota Department of Education	$2.00
_____	Assessment Theodore E. Andrews, Editor	$2.00

To order publications from the National Dissemination Center, check the titles you require, indicating number of copies needed, and return this form to the **National Dissemination Center, Syracuse University, 123 Huntington Hall, Syracuse, New York 13210.**

All orders not accompanied by a formal purchase order must include a check or money order made payable to the **National Dissemination Center.**

MAILING AND HANDLING CHARGE:

U.S. and Canada:

for one publication	$0.60
for more than one publication	$1.00

FOREIGN:

surface mail	$0.50 publication
air mail	$2.00/ publication

Payment enclosed____ Purchase Order No.____ Amount $_____

NAME _____

ADDRESS _____

_____ ZIP_____